Exhibit 7

7/20/23, 11:38 PM

Case 1:23-cv-11775-MJJ Document 27 Filed 08/08/23 Page 2 of 9
[110] Data Falsificada (Part 2): "My Class Year Is Harvard" – Data Colada

Thinking about evidence, and vice versa

HOME   TABLE OF CONTENTS   FEEDBACK POLICY   SEMINAR   ABOUT

[110] Data Falsificada (Part 2): "My Class Year Is Harvard"

Posted on June 20, 2023 by Uri, Joe, & Leif

This is the second in a four-part series of posts detailing evidence of fraud in four academic papers co-authored by Harvard Business School Professor Francesca Gino. It is worth reiterating that to the best of our knowledge, none of Gino's co-authors carried out or assisted with the data collection for the studies in this series.

A message for the 148 co-authors of Francesca Gino who may be reading these posts. A small subset of co-authors (from multiple universities, including Harvard) is organizing an open and collegial effort to provide a centralized and standardized report on the origin of data for all studies published in papers co-authored by Gino. One of the primary goals of this effort is to protect the work and careers of young researchers by promptly identifying published studies in papers co-authored by Gino, but for which she was not in charge of the data collection or analyses. Within the next 2-4 weeks, you will all be receiving an email inviting you to participate. We do not have more information about this effort at this time.

Part 2: My Class Year Is Harvard

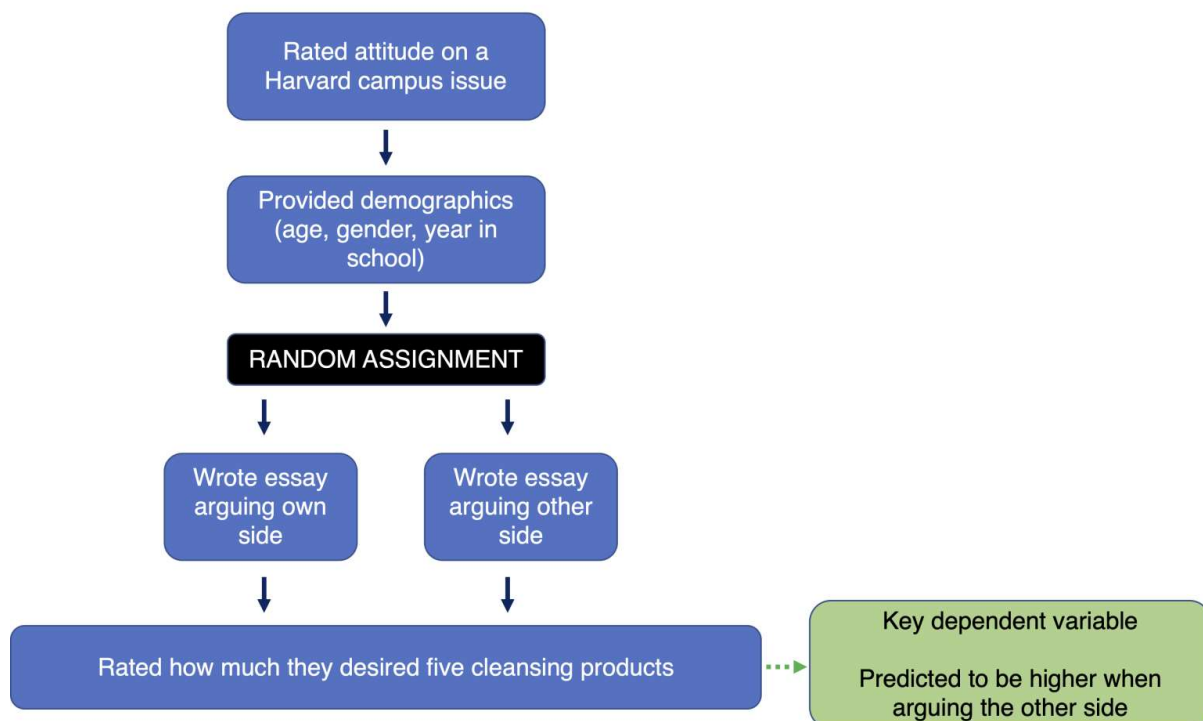Gino, Kouchaki, & Galinsky (2015), Study 4
"The Moral Virtue of Authenticity: How Inauthenticity Produces Feelings of Immorality and Impurity", *Psychological Science*

In this paper, the authors presented five studies suggesting that experiencing inauthenticity leads people to feel more immoral and impure (.htm). Here we focus on Study 4, which was run at Harvard University.

Harvard students (N = 491) completed an online survey. They were first asked to express an opinion about a Harvard campus issue and to provide some demographic information. Then they were asked to write an essay about that issue that either argued in favor of their own side or against their own side [1].

After writing the essay, participants rated how desirable they found five cleansing products to be (1 = Completely undesirable; 7 = Completely desirable). The authors predicted that writing against your own side would make participants feel dirty/impure, which would increase their desire for cleansing products.

Gino, Kouchaki, & Galinsky (2015) Study 4 Procedure
(Online Study with Harvard Student Participants)

Rated attitude on a Harvard campus issue
↓
Provided demographics (age, gender, year in school)
↓
RANDOM ASSIGNMENT
↓ ↓
Wrote essay arguing own side / Wrote essay arguing other side
↓ ↓
Rated how much they desired five cleansing products ┄▶ Key dependent variable — Predicted to be higher when arguing the other side

Results

Consistent with the authors' hypotheses, participants desired cleansing products more when arguing against (*M* = 4.26, *SD* = 1.48) vs. in favor of (*M* = 3.72, *SD* = 1.33) their own side, $t(488) = 3.93$, $p < .0001$.

The Anomaly: Strange Demographic Responses

As mentioned above, students in this study were asked to report their demographics. Here is a screenshot of the posted original materials, indicating exactly what they were asked and how:

4. Your age: _____

5. Your gender
   • Male
   • Female
   • Other (please indicate)

6. Year in School: _____

We retrieved the data from the OSF (https://osf.io/sd76g), where Gino (or someone using her credentials) posted it in 2015. The anomaly in this dataset involves how some students answered Question #6: "Year in School."
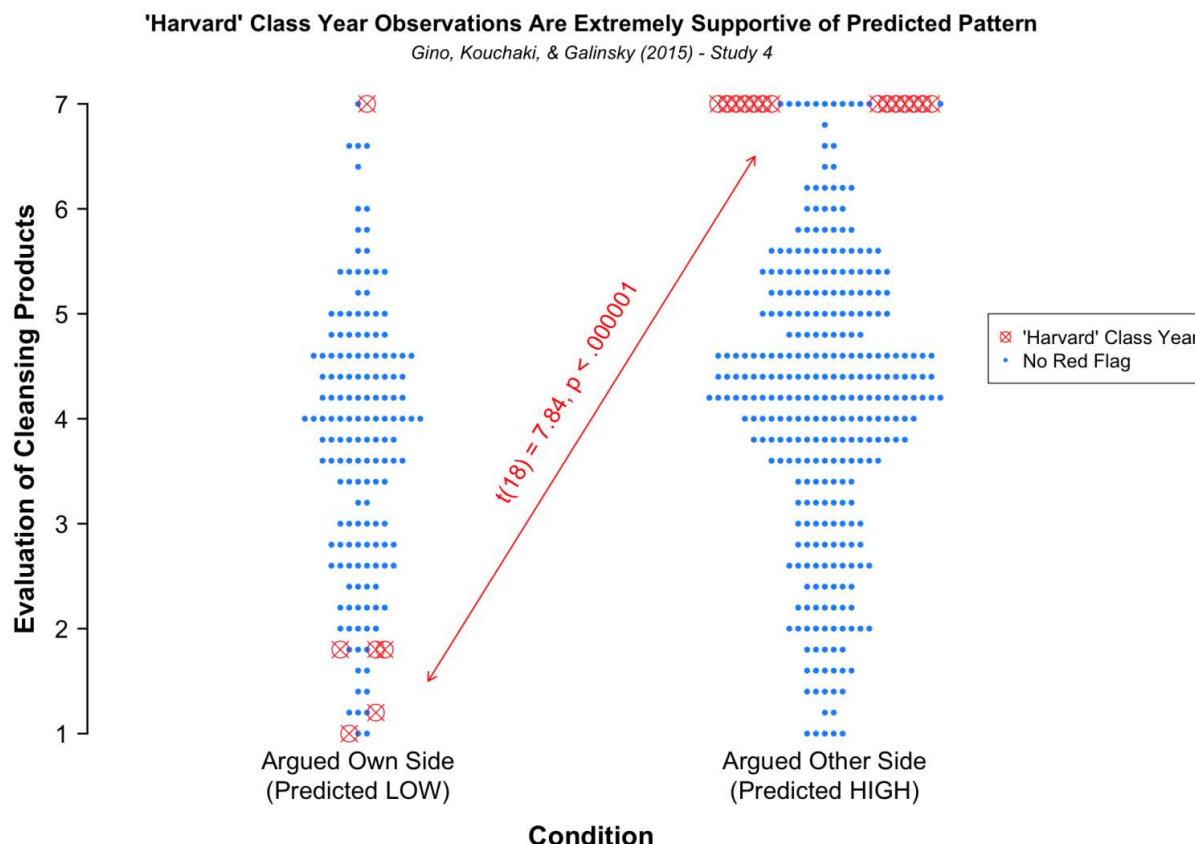
The screenshot below shows a portion of the dataset. In the "yearSchool" column, you can see that students approached this "Year in School" question in a number of different ways. For example, a junior might have written "junior", or "2016" or "class of 2016" or "3" (to signify that they are in their third year). All of these responses are reasonable.

A less reasonable response is "Harvard", an incorrect answer to the question. It is difficult to imagine many students independently making this highly idiosyncratic mistake. Nevertheless, the data file indicates that 20 students did so. Moreover, and adding to the peculiarity, those students' responses are all within 35 rows (450 through 484) of each other in the posted dataset:

	A	instr	college_student	inFavor_inclusion_diffRating_in_Qguide	strongOpin	age	male	gender_TEXT	yearSchool
448	447	1	1	1	5	22	1		Senior
449	448	1	1	1	5	21	1		Senior
450	449	1	1	1	7	23	0		harvard
451	450	1	1	1	4	21	0		2015
452	451	1	1	1	7	20	1		Junior
453	452	1	1	1	7	18	0		Sophomore
454	453	1	1	0	7	25	0		Harvard
455	454	1	1	0	7	25	0		Harvard
456	455	1	1	1	7	22	1		Harvard
457	456	1	1	0	7	24	0		Harvard
458	457	1	1	1	7	22	0		Harvard
459	458	1	1	0	7	25	0		Harvard
460	459	1	1	1	7	23	1		Harvard
461	460	1	1	0	7	25	0		Harvard
462	461	1	1	0	6	25	1		4
463	462	1	1	0	7	24	0		Harvard
464	463	1	1	1	5	23	0		1
465	464	1	1	1	4	19	0		Sophomore
466	465	1	1	1	6	28	1		5
467	466	1	1	1	6	22	1		Senior
468	467	1	1	1	6	20	0		Junior
469	468	1	1	1	5	23	1		2015
470	469	1	1	1	6	22	1		Senior
471	470	1	1	1	6	22	0		2015/Senior
472	471	1	1	1	6	36	1		2010
473	472	1	1	1	7	25	0		Harvard
474	473	1	1	0	5	25	0		Harvard
475	474	1	1	1	7	22	1		Harvard
476	475	1	1	1	7	23	1		Harvard
477	476	1	1	0	7	25	0		Harvard
478	477	1	1	0	7	26	1		Harvard
479	478	1	1	1	6	20	0		2013
480	479	1	1	0	6	21	0		2012
481	480	1	1	1	7	24	1		Harvard
482	481	1	1	1	7	27	0		Harvard
483	482	1	1	1	7	25	1		Harvard
484	483	1	1	1	7	27	0		Harvard
485	484	1	1	1	7	26	1		4
486	485	1	1	0	6	22	0		2012
487	486	1	1	1	6	20	1		2013
488	487	1	1	0	6	20	0		2013

On its own, this is a bright red flag. There is simply no reason why so many students would provide such a strange answer to this question. It suggests that someone may have tampered with the data in a way that left so many "My class year is Harvard" responses.

If these peculiar observations were indeed tampered with, then we should see that students who answered "Harvard" were especially likely to confirm the authors' hypothesis. The plot below depicts the key dependent variable – participants' average ratings of how much they desired five cleansing products – for every observation. Every "normal" observation is represented as a blue dot, whereas the 20 "Harvard" observations are represented as red X's:

'Harvard' Class Year Observations Are Extremely Supportive of Predicted Pattern
Gino, Kouchaki, & Galinsky (2015) - Study 4

You can see that in the argued-other-side conditions every "Harvard" observation had the highest possible average value (i.e., a 7.0). Conversely, in the argued-own-side condition every "Harvard" observation was associated with a low value, except for one [2]

The difference between the two conditions for just these 20 observations is highly significant ($p < .000001$). Moreover, the effect for the 'Harvard' observations is significantly larger than the effect for the non-Harvard observations ($p < .000001$).

This strongly suggests that these 'Harvard' observations were altered to produce the desired effect. And if these observations were altered, then it is reasonable to suspect that other observations were altered as well.

Indeed, we should note that while for all four studies covered in this series we found evidence of data tampering, we do not believe (in the least) that we've identified all of the tampering that happened within these studies. Without access to the original (un-tampered) data files – files we believe Harvard had access to – we can only identify instances when the data tamperer slipped up, forgetting to re-sort here, making a copy-paste error there. There is no reason (at all) to expect that when a data tamperer makes a mistake when changing one thing in a database, that she makes the same mistake when changing all things in that database.

When the series is over we will post all code, materials, and data on a single ResearchBox. In the meantime:

https://datacolada.org/appendix/110/

Author feedback.

Our policy is to solicit feedback from authors whose work discuss. We did not do so this time, given (1) the nature of the post, (2) that the claims made here were presumably vetted by Harvard University, (3) that the articles we cast doubt on have already had retraction requests issued, and (4) that discussions of these issues were already surfacing on social media and by some science journalists, without their having these facts, making a traditional multi-week back-and-forth with authors self-defeating.

Subscribe to Blog via Email

Enter your email address to subscribe to this blog and receive notifications of new posts by email.

Email Address

Subscribe

Footnotes.

1. There were two argue-against-their-own-side conditions that differed in whether they were "forced" to write against their own side, or were strongly encouraged to write against their own side. Those condition differences don't matter much, so we collapse across them here. [↵]

2. In a previous version of this post, we incorrectly said that this one inconsistent observation was the only one associated with a lowercase 'harvard'. That is not the case, as the lowercase 'harvard' observation is in the Argued Other Side condition. We thank Steve Haroz for taking the time to re-analyze our data, find the error, and notify us. We corrected this error at 11:34 am ET on June 20, 2023. [↵]

GET COLADA EMAIL ALERTS.

you@email.com

Subscribe

Join 6,578 other subscribers

SOCIAL MEDIA

🐦 We tweet new posts: @DataColada
Ⓜ️And mastopost'em: @DataColada@mas.to
f And link to them on our Facebook page

RECENT POSTS

[112] Data Falsificada (Part 4): "Forgetting The Words"

[111] Data Falsificada (Part 3): "The Cheaters Are Out of Order"

[110] Data Falsificada (Part 2): "My Class Year Is Harvard"

[109] Data Falsificada (Part 1): "Clusterfake"

[108] MRAN is Dead, long live GRAN

GET BLOGPOST EMAIL ALERTS

POSTS ON SIMILAR TOPICS

your@email.com

Fake data

Submit

Join 6,578 other subscribers

TWEETER & FACEBOOK

We tweet new posts: @DataColada

And link to them on our Facebook page

[112] Data Falsificada (Part 4): "Forgetting The Words"

[111] Data Falsificada (Part 3): "The Cheaters Are Out of Order"

[110] Data Falsificada (Part 2): "My Class Year Is Harvard"

[109] Data Falsificada (Part 1): "Clusterfake"

[98] Evidence of Fraud in an Influential Field Experiment About Dishonesty

[77] Number-Bunching: A New Tool for Forensic Data Analysis

[74] In Press at Psychological Science: A New 'Nudge' Supported by Implausible Data

[40] Reducing Fraud in Science

[21] Fake-Data Colada: Excessive Linearity

[19] Fake Data: Mendel vs. Stapel

SEARCH

Search …

Search